AD_____

Award Number: W81XWH-11-1-0526

TITLE: Collaborative Model for Acceleration of Individualized Therapy of Colon Cancer

PRINCIPAL INVESTIGATOR: S. Gail Eckhardt, M.D.

CONTRACTING ORGANIZATION: UNIVERSITY OF COLORADO
AURORA CO 80045-2505

REPORT DATE: October 2012

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
10/1/2012	Annual	15 Sept 2011 – 14 Sept 2012

4. TITLE AND SUBTITLE

Collaborative Model for Acceleration of Individualized Therapy of Colon Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0526

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
S. Gail Eckhardt, M.D., Aik Choon Tan, Ph.D. and Todd M. Pitts, M.S.
Betty Diamond

E-Mail: gail.eckhardt@ucdenver.edu;aikchoon.tan@ucdenver.edu;todd.pitts@ucdenver.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

UNIVERSITY OF COLORADO

AURORA CO 80045-2505

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command

Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Colorectal cancer (CRC) represents a major health burden, and is the third leading cause of cancer deaths in the U.S. In the past decade, the median survival among patients with metastatic CRC has significantly improved, primarily due the development of active chemotherapeutic regimens that include biological agents. However, despite this success, patients soon run out of therapeutic options and receive salvage therapy that results in only a few weeks of disease stability. We have proposed to employ a team science, systems biology based approach to rapidly identify novel anti-cancer agents and individualize therapeutic strategies in preclinical CRC models. In this Year 1 Progress report, we will present the tasks and key accomplishments achieved within this period of time. In brief, we have completed in vitro testing on a large panel of CRC cell lines for six novel anti-cancer agents. We have completed baseline gene expression profiling of our CRC cell lines panel and patient-derived CRC tumor explant models by high-throughput RNA sequencing approach. We have initiated the in vivo cell line derived xenograft models to test the efficacy of these novel anti-cancer agents and in the process of determining the down stream effectors of these targets by immunoblotting assays. Our research findings for RNA-seq analaysis will be presented at the 24th EORTC-NCI-AACR Symposium on Molecular Targets and Cancer Therapeutics, Dublin, Ireland (November 6-9, 2012). In summary, we have accomplished all the tasks that we proposed in year 1.

15. SUBJECT TERMS
Colorectal cancer, novel anti-cancer agents, bioinformatics, next-generation sequencing, predictive biomarkers, individualized medicine

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	14	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Progress Report for Year 1.

Introduction: Colorectal cancer (CRC) represents a major health burden, and is the third leading cause of cancer deaths in the U.S. In the past decade, the median survival among patients with metastatic CRC has significantly improved, primarily due the development of active chemotherapeutic regimens that include biological agents. However, despite this success, patients soon run out of therapeutic options and receive salvage therapy that results in only a few weeks of disease stability. This is particularly true for a subset of patients that have a mutation in the KRAS gene, since it has been shown that one of these new treatments is not effective for them. Therefore, new agents are needed that can stabilize disease and hopefully prolong life in patients with CRC. One of the lessons learned in CRC, in fact, in patients with the KRAS mutation in their tumor, is the importance of not only developing new effective drugs, but also developing ways to select patients for those treatments. Unfortunately the lack of such strategies is what led to thousands of CRC patients with KRAS mutations being treated with epidermal growth factor receptor (EGFR) inhibitors at considerable toxicity and no benefit, when it was discovered that tumors with this mutation did not respond to these drugs. This new area of patient selection, or individualized therapy, is based upon a robust set of research tools in the field of bioinformatics. Therefore, successful research teams are comprised of clinicians, who treat patients with cancer, and bioinformaticians, that are able to synthesize large sets of data and look for patterns of response or resistance to a particular new drug. Such a team has been assembled for this proposal. Thus, the overall goal of this Idea Award is enhance the efficiency and speed of developing novel and individualized therapy for patients with KRAS mutant colorectal cancer (CRC) using a comprehensive bioinformatics approach and novel preclinical models of human CRC. This proposal has the potential of providing novel, individualized therapeutic strategies for CRC patients with KRAS mutations that are poised for clinical testing at the completion of this work (3 years). The yield will be highly relevant, as new drug development will not only be jump-started by this proposal but agents to be tested clinically will be tailored for specific populations of patients with CRC, thereby potentially conferring greater clinical benefit. In this progress report, we will describe our research achievements and outcomes for **Year 1**.

Aim 1. To develop predictive classifiers for 3 novel agents using preclinical models of colorectal cancer (CRC).
We have selected the following novel agents for initial screens using preclinical models of colorectal cancer.

Table 1: Six novel anti-cancer agents selected in this study.

Agents	**Targets**	**Company**	**Clinical Developmental Phase**
MLN8237 (alisertib)	Aurora Kinase A (AURKA)	Millennium Pharmaceuticals/Takeda	Phase I
TAK733	Dual specificity mitogen-activated protein kinase kinase 1 (MAP2K1)	Millennium Pharmaceuticals/Takeda	Phase I
TAK960	Polo-like Kinase 1 (PLK1)	Millennium Pharmaceuticals/Takeda	Phase I
MLN0128	TORC1/TORC2	Millennium Pharmaceuticals/Takeda	Phase I
ENMD2076	Aurora Kinase A (AURKA) and Angiogenic Kinase (KDR)	EntreMed	Phase I/II
PF-04691502	Phosphatidylinositol 3-Kinase (PIK3CA) and mammalian Target of Rapamycin (mTOR)	Pfizer	Phase I

Task 1: *In vitro* cell line exposure (Months 1-12, Dr. Eckhardt).

To evaluate the sensitivity of CRC cell lines to MLN8237, TAK960, TAK733, ENMD2076, PF-04691502 and MLN0128, a panel of CRC cell lines were exposed to increasing concentrations of these novel anti-cancer agents and assessed for proliferation using an SRB assay as previously described (Skehan et al 1990; Pitts et al 2010). As depicted in **Figure 1** there was a broad range of sensitivity of the CRC cell lines to these anti-cancer agents, *indicating that patient selection is needed*.

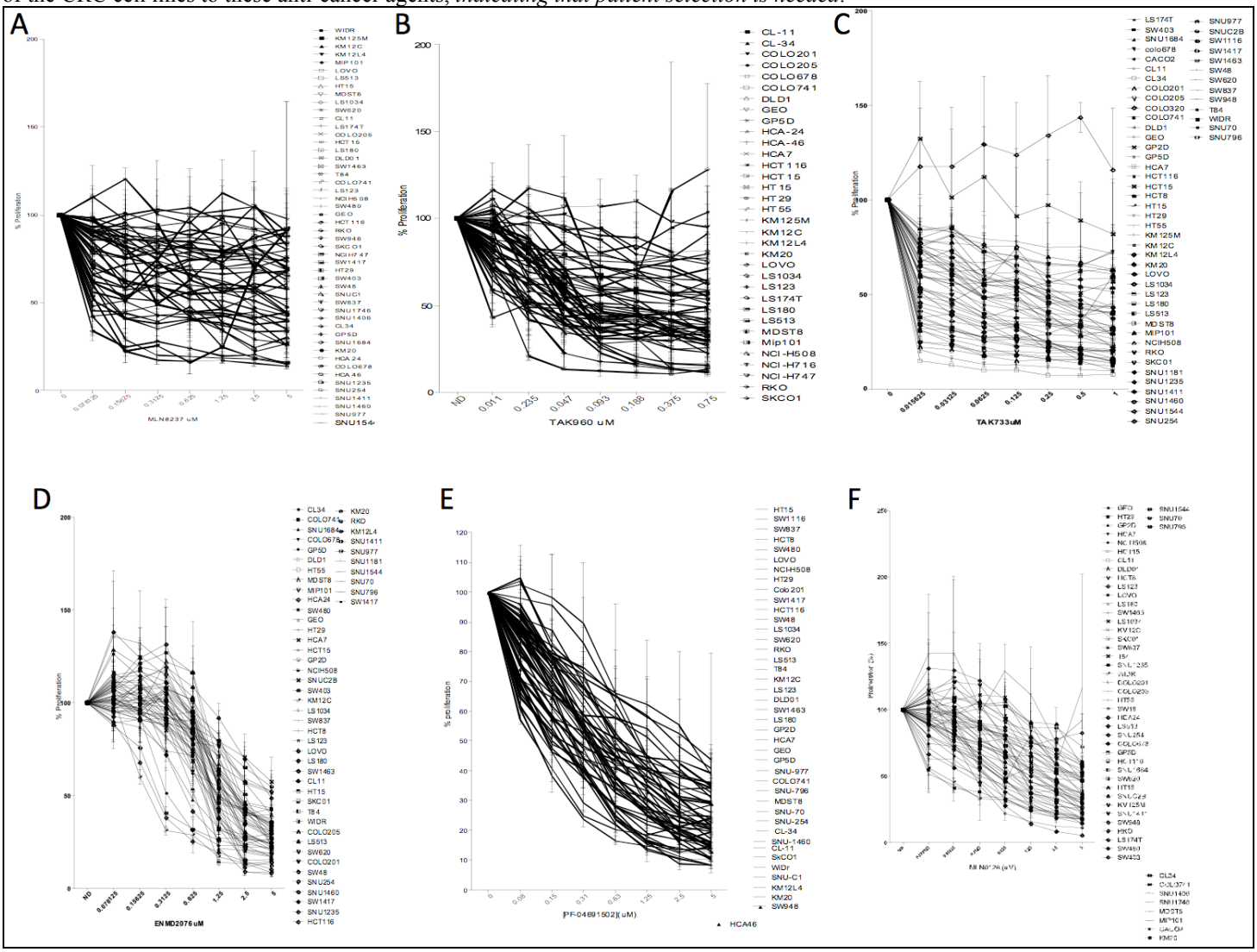

Figure 1: A panel of CRC cell lines were exposed to increasing concentrations of MLN8237 (A), TAK960 (B), TAK733 (C), ENMD2076 (D), PF-04691502 (E), and MLN0128 (F).

Task 2: *In vivo* cell line xenograft treatment (Months 6-18, Dr. Eckhardt).
To determine the *in vivo* inhibition, we have performed treatment using these anti-cancer agents on cell lines derived xenografts as previously described (Pitts et al 2010). We have treated three CRC cell line xenografts with MLN8237 (**Figure 2**), TAK960 (**Figure 3**), TAK733 (**Figure 4**), ENMD2076 (**Figure 5**), and two with PF-04691502 (**Figure 6**). We are in the process finishing this task (*Months 12-18*) by injecting more mice with CRC cell lines and treating with the compounds listed. As anticipated, there is also a diversity of responses to these agents *in vivo*.

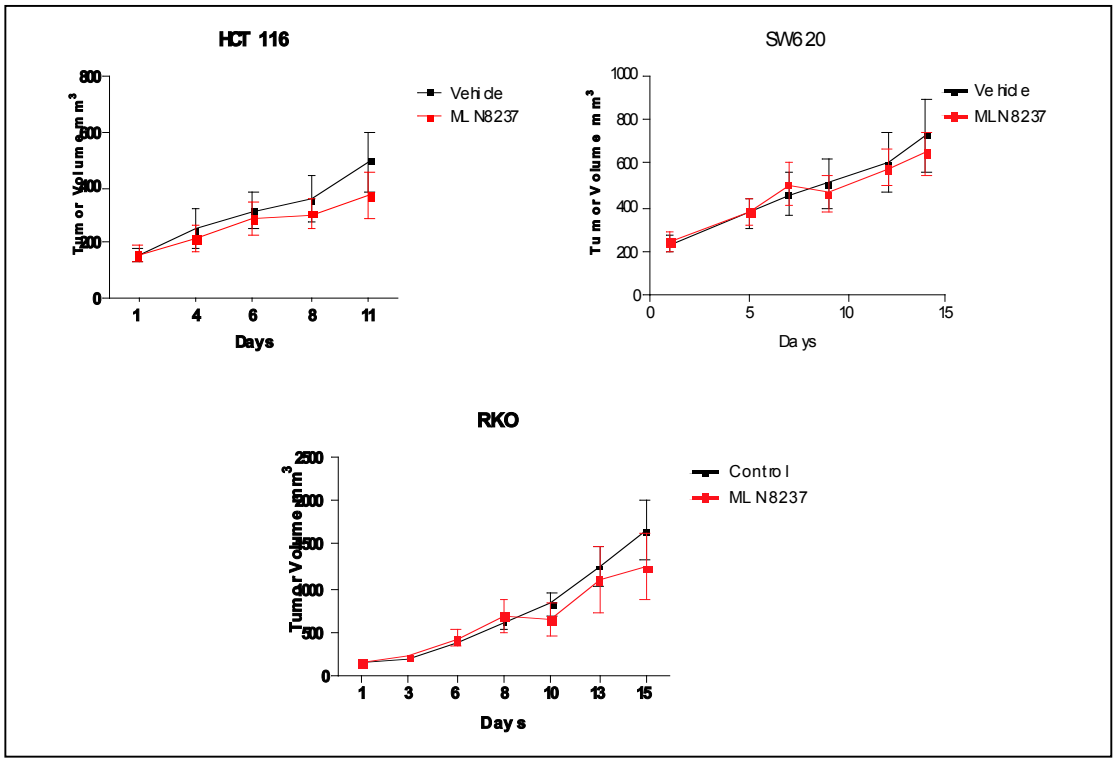

Figure 2: *In vivo* cell lines treated with MLN8237.

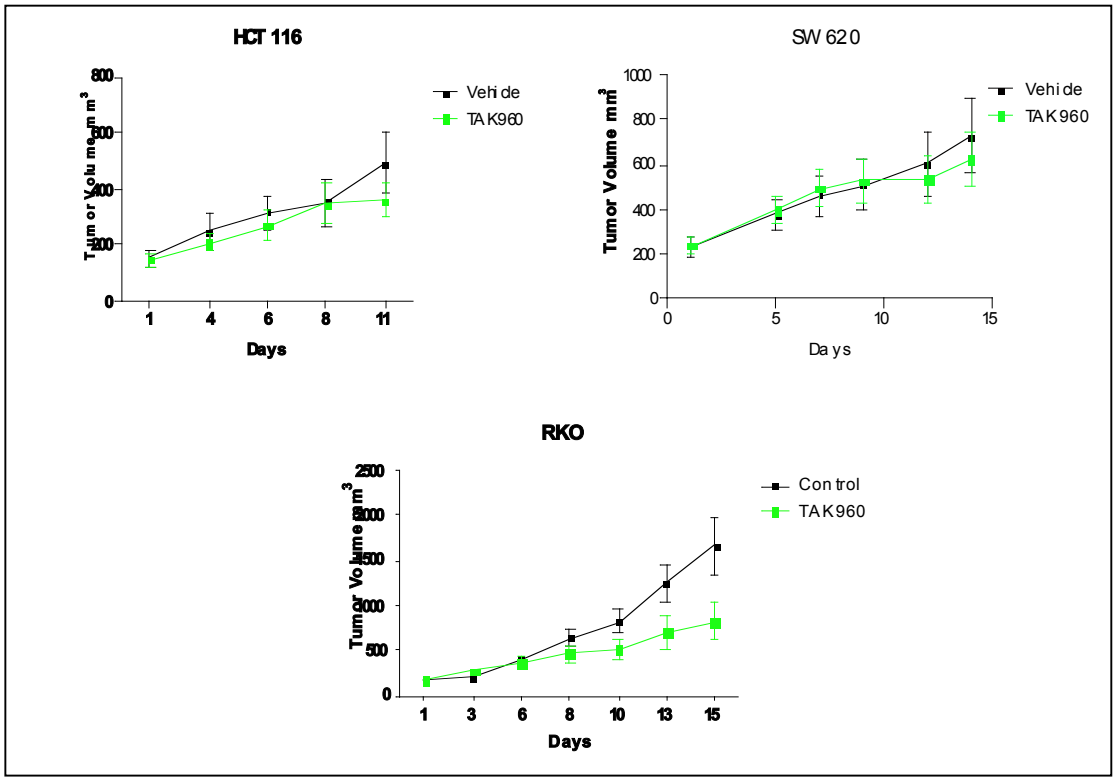

Figure 3: *In vivo* cell lines treated with TAK960.

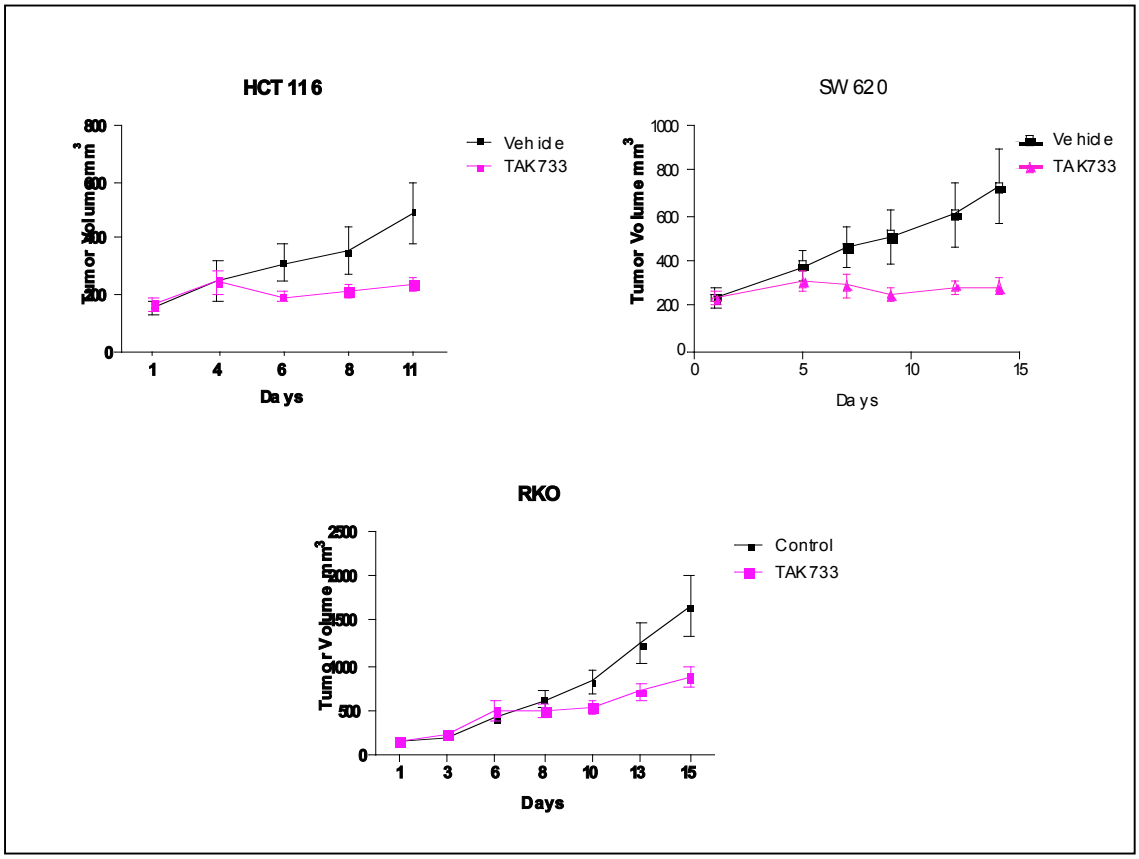

Figure 4: *In vivo* cell lines treated with TAK733.

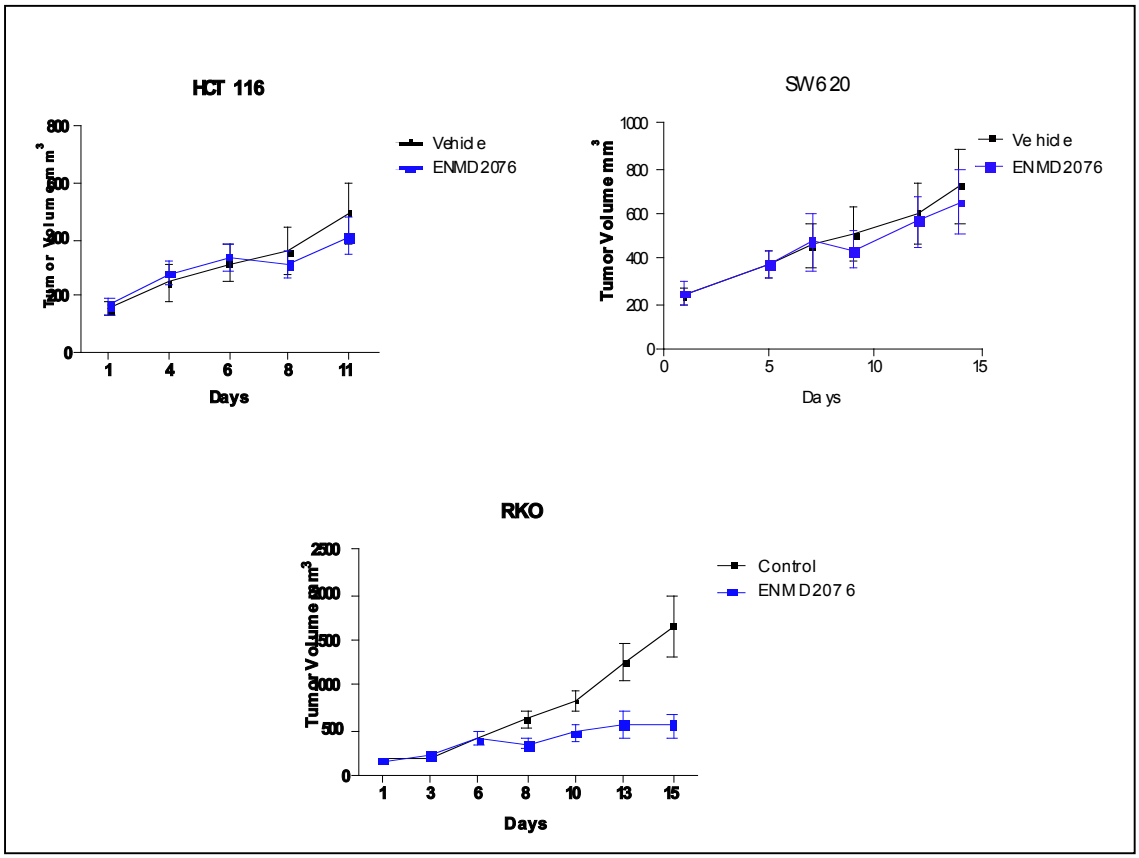

Figure 5: *In vivo* cell lines treated with ENMD2076.

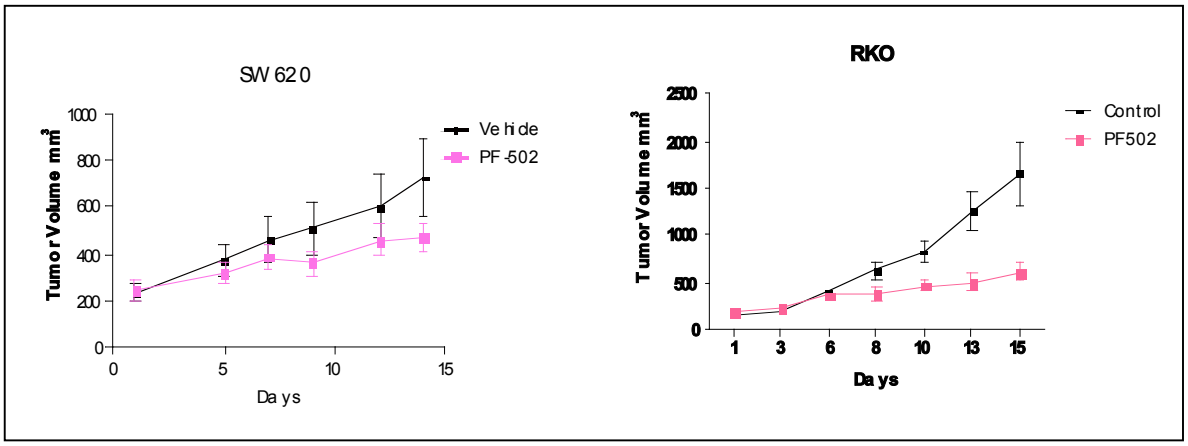

Figure 6: *In vivo* cell lines treated with PF-04691502.

Task 3: Immunoblotting for relevant downstream effectors (Months 6-18, Dr. Eckhardt).
To access the inhibition of these anti-cancer agents in the cancer cells, we have performed immunoblotting for relevant downstream effectors of these targets. As depicted in **Figure 7**, six CRC cell lines were exposed to MLN8237 or TAK733 for 24 hours. Protein was extracted and westerns were performed to look at downstream effectors. We are in the process of exposing more CRC cell lines to the other compounds and performing westerns, which will be completed *Months 12-18*. These results demonstrate that although downstream effector modulation may document pharmacodynamic effects, they are not sufficient for patient selection.

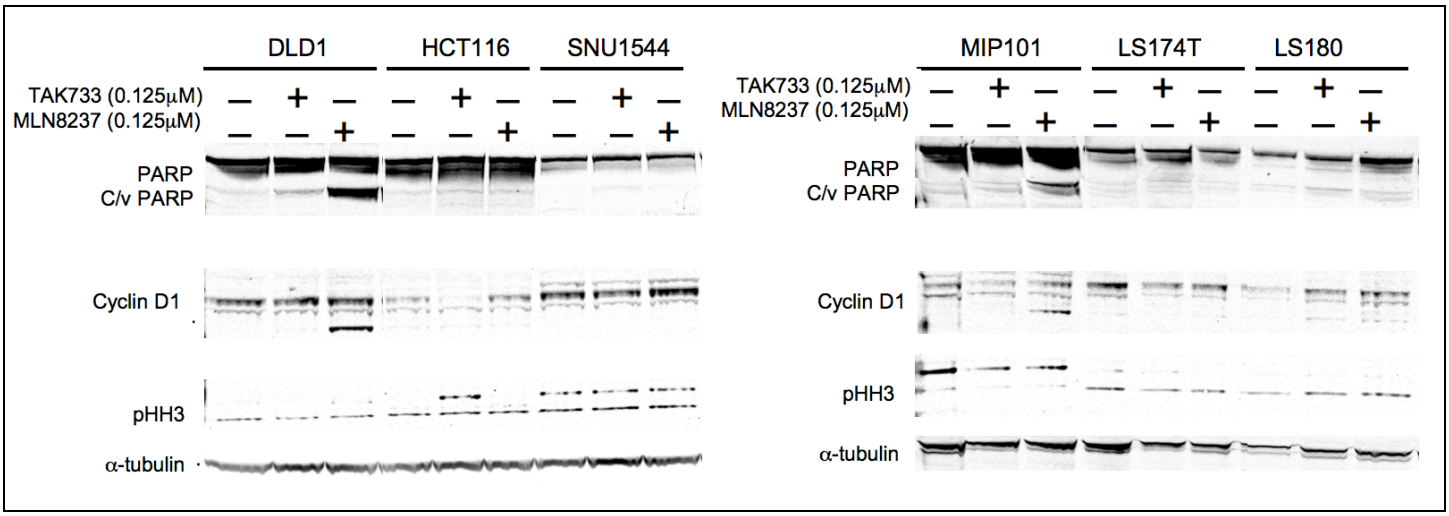

Figure 7: Immunoblotting for relevant downstream effectors of MLN8237 or TAK733 in six CRC cell lines.

Task 4: Perform transcriptome sequencing (RNA-Seq) on CRC cell lines (*in vitro* and xenografts) (Months 1-18, Dr. Tan).
Total RNAs were extracted from the cancer cells or tumor tissues using Trizol (Invitrogen, Carlsbad, CA). Libraries were constructed using 1μg total RNA following Illumina TruSeq RNA Sample Preparation v2 Guide. The poly-A containing mRNA molecules were purified using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were converted into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments then were subjected to an end repair process, the addition of a single "A" base, and ligation of the adapters. The products were purified and enriched using PCR to create the final cDNA library. The cDNA library was validated on the Agilent 2100 Bioanalyzer using DNA-1000 chip. Cluster generation was performed on the Illumina cBot using a Single Read Flow Cell with a Single Read cBot reagent plate (TruSeq SR Cluster Kit v3-cBot-HS). Sequencing of the clustered flow cell was performed on the Illumina HiSeq 2000 using TruSeq SBS v3 reagents. We used the Illumina HiSeq2000 as this is the latest machine with higher sequencing throughput and cheaper for sequencing cost. Utilizing the latest HiSeq2000 machine, we were able to multiplex 3 samples per lane, sequence with single end 100 cycles (1x100bp) and achived ~40 million reads per sample. The number of cycles for each read is also programmed into the machine before the run begins. Sequencing images were generated through the sequencing platform (Illumina HiSeq 2000). The raw data were analyzed in four steps: image analysis, base calling, sequence alignment, and variant analysis and counting. An additional step was required to convert the base call files (.bcl) into *_qseq.txt files. For multiplexed lanes/samples, a de-multiplexing step is performed before the alignment step.

Task 5: Bioinformatics analysis of RNA-Seq data (Months 12-18, Dr. Tan).

High-throughput mRNA sequencing (RNAseq) of each sample was obtained from the Illumina HiSeq2000. On average, we obtained about 60 million (coverage ranged from 30 to 90 million reads) single-end 100bp sequencing reads per sample. To analyze the RNAseq data, the reads were mapped against the human genome using the BiNGS! (Bioinformatics for Next Generation Sequencing) pipeline. In our pipeline, we have optimized the parameters for mapping using Tophat (Trapnell et al 2009) and cufflinks (Trapnell et al 2010). The first step of the BiNGS! pipeline is mapping the reads against the reference genome. Here, we used the NCBI reference annotation (build 37.2) as a guide, and allowing 3 mismatches for the initial alignment and 2 mismatches per segment with 25 bp segments using Tophat (version 1.3.2). On average, 92% (ranging from 71% to 95%) of the reads aligned to the human genome. Next, the workflow employed Cufflinks (version 1.3.0) to assemble the transcripts using the RefSeq annotation as the guide, but allowing for novel isoform discovery in each sample. Isoforms were ignored if the number of supporting reads was less than 30 and if the isoform fraction was less than 10% for the gene. The data were fragment bias corrected, multi-read corrected, and normalized by the total number of reads. On average, the sequences can be mapped to 20,221 known genes (ranging from 18,213 to 21,448 genes). The transcript assemblies for each sample were merged using cuffmerge. To estimate the transcript expressions of individual sample, we computed the FPKM values of the transcripts by rerunning Cufflinks again using the merged assembly as the guide. The final output of this analysis step is a P x N matrix, where P is the number of samples and N is the number of transcripts, respectively. Gene expression for individual sample is estimated by summing the FPKM values of multiple transcripts that represent the same gene. Subsequent data analyses of RNAseq will be performed on this matrix. **Table 1** summarizes the RNA-seq results for the 55 colorectal cancer cell lines.

Table 1: RNA-seq results for the colorectal cancer cell lines.

Colorectal Cancer Cell Lines	Number of Reads	Number of mappable reads (one or more hits)	Mappability (%)	Number of known genes
CACO2	40,904,569	37,993,175	92.9%	20,297
CL11	78,658,444	73,181,227	93.0%	21,151
CL34	69,389,421	63,580,698	91.6%	20,306
COLO201	57,177,587	49,067,703	85.8%	20,264
COLO205	63,560,880	58,898,846	92.7%	19,481
COLO678	57,230,643	46,088,019	80.5%	19,564
COLO741	65,015,416	61,020,161	93.9%	20,891
DLD1	52,467,265	47,719,907	91.0%	20,425
GEO	69,811,802	64,957,702	93.0%	19,448
GP2D	59,787,549	56,112,822	93.9%	19,517
GP5D	63,532,093	59,190,555	93.2%	19,886
HCA24	61,157,777	57,272,556	93.6%	19,967
HCA46	54,698,858	50,698,239	92.7%	21,090
HCA7	66,112,001	62,614,913	94.7%	20,745
HCT116	60,005,963	56,446,384	94.1%	20,642
HCT15	62,871,876	58,963,133	93.8%	20,050
HCT8	64,659,575	52,469,328	81.1%	19,917
HT29	57,066,621	54,126,579	94.8%	19,527
HT55	77,945,763	73,055,027	93.7%	20,954
KM12C	61,028,210	55,338,944	90.7%	20,375
LOVO	54,471,343	50,967,636	93.6%	19,209
LS1034	39,468,606	36,537,823	92.6%	19,862
LS123	60,903,455	56,433,627	92.7%	21,448
LS174T	69,755,740	64,985,999	93.2%	21,088
LS180	62,836,493	50,187,948	79.9%	19,926
LS513	51,459,126	47,957,782	93.2%	18,878
MDST8	70,862,640	66,338,334	93.6%	20,438
MIP101	60,620,723	56,951,570	93.9%	20,028
NCIH508	65,734,534	61,463,550	93.5%	19,516
NCLH747	63,594,163	59,597,210	93.7%	21,260
RKO	52,278,860	49,153,319	94.0%	18,213
SKCO1	58,920,742	54,946,092	93.3%	20,540
SNU1181	65,188,766	60,868,362	93.4%	21,092
SNU1235	61,625,438	57,337,118	93.0%	20,768
SNU1406	75,515,317	70,595,915	93.5%	20,195
SNU1411	70,702,006	66,239,041	93.7%	20,873
SNU1460	30,073,092	28,239,886	93.9%	19,420
SNU1544	67,551,529	62,315,207	92.2%	20,249
SNU1684	61,395,914	57,711,989	94.0%	20,392

SNU1746	55,470,908	51,696,812	93.2%	18,892
SNU254	47,062,873	43,675,944	92.8%	18,502
SNU70	90,904,920	85,188,933	93.7%	20,892
SNU796	48,259,764	45,230,481	93.7%	20,775
SNU977	51,197,362	48,236,308	94.2%	20,271
SNUC2B	49,697,829	35,056,907	70.5%	20,848
SW1116	58,638,376	54,786,346	93.4%	19,935
SW1463	61,655,600	57,823,324	93.8%	21,219
SW403	57,484,840	53,533,120	93.1%	20,229
SW48	49,807,298	46,140,124	92.6%	19,396
SW480	58,120,265	54,776,367	94.2%	21,157
SW480	49,838,558	46,930,672	94.2%	21,002
SW620	69,770,696	65,954,353	94.5%	20,663
SW837	88,060,727	82,396,936	93.6%	21,007
SW948	60,917,538	57,306,844	94.1%	20,200
WIDR	39,216,201	36,609,136	93.4%	19,264

Similar to the colorectal cancer cell lines, we performed RNA-seq for 25 colorectal cancer explants. Using the same RNAseq protocol, we sequenced the tumor samples on single-end 100 bp with Illumina HiSeq2000, multiplexing 3 samples per lane. On average, we obtained about 57 million (coverage ranged from 43 to 83 million reads) single-end 100bp sequencing reads per sample. To analyze the RNAseq data, the reads were mapped against the human genome using the BiNGS! workflow. We used the NCBI reference annotation (build 37.2) as a guide, and allowing 3 mismatches for the initial alignment and 2 mismatches per segment with 25 bp segments using Tophat (version 1.3.2). On average, 84% (ranging from 68% to 92%) of the reads aligned to the human genome. Next, the workflow employed Cufflinks (version 1.3.0) to assemble the transcripts using the RefSeq annotation as the guide, but allowing for novel isoform discovery in each sample. Isoforms were ignored if the number of supporting reads was less than 30 and if the isoform fraction was less than 10% for the gene. The data were fragment bias corrected, multi-read corrected, and normalized by the total number of reads. On average, the sequences can be mapped to 19,355 known genes (ranging from 17,481 to 21,519 genes). The transcript assemblies for each sample were merged using cuffmerge. To estimate the transcript expressions of individual sample, we computed the FPKM values of the transcripts by rerunning Cufflinks again using the merged assembly as the guide. The final output of this analysis step is a P x N matrix, where P is the number of samples and N is the number of transcripts, respectively. Gene expression for individual sample is estimated by summing the FPKM values of multiple transcripts that represent the same gene. Subsequent data analyses of RNAseq will be performed on this matrix. **Table 2** summarizes the RNA-seq results for the 25 colorectal cancer explants.

Table 2: RNA-seq for colorectal cancer explants.

Sample	Number of reads	Number of mappable reads (one or more hits)	Mappability (%)	Known genes
CRC001	70,493,980	61,441,833	87.2%	18,132
CRC006	51,874,201	41,893,143	80.8%	19,861
CRC007	42,969,115	34,477,137	80.2%	18,116
CRC010	45,120,688	36,431,454	80.7%	19,803
CRC012	60,650,098	54,174,567	89.3%	18,274
CRC020	49,709,595	44,847,503	90.2%	18,446
CRC021	58,415,554	39,711,426	68.0%	18,042
CRC026	49,217,390	44,333,105	90.1%	19,457
CRC027	49,979,478	46,055,247	92.1%	19,957
CRC034	51,287,241	38,124,482	74.3%	19,990
CRC035	70,168,636	60,658,624	86.4%	20,162
CRC036	83,826,375	76,843,848	91.7%	19,594
CRC040	48,857,900	41,352,127	84.6%	19,243
CRC047	70,462,203	53,990,167	76.6%	18,899
CRC052	43,106,395	32,441,529	75.3%	17,481
CRC065	67,185,367	59,084,004	87.9%	21,519
CRC098	66,526,059	53,099,856	79.8%	20,574
CRC099	54,564,275	42,922,695	78.7%	19,735
CRC101	56,983,496	45,159,010	79.2%	19,559
CRC102	63,805,948	57,267,897	89.8%	20,166
CRC106	62,919,597	43,270,369	68.8%	17,898
CRC108	69,409,511	61,925,543	89.2%	20,100
CRC114	43,453,943	37,545,373	86.4%	18,918
CRC125	48,899,497	43,588,069	89.1%	20,132
CRC138	46,308,170	41,614,273	89.9%	19,815

To explore the unmappable reads from the colorectal cancer explants against human genome, we mapped these remaining reads against the mouse genome (NCBI reference annotation build 37.2) using the same BiNGS! pipeline. On average, 5.9% of these remaining reads were mapped to mouse genome, indicating that the tumor samples that we extracted from the explants are highly enriched with human cancer cells. **Table 3** summarizes the mapping results.

Table 3: Mapping results of the CRC explants against human and mouse genomes.

Sample	% of reads aligned to human genome	% of reads aligned to mouse genome	Total Mappability (%)
CRC001	87.2%	4.3%	91.4%
CRC006	80.8%	14.1%	94.9%
CRC007	80.2%	7.4%	87.6%
CRC010	80.7%	13.8%	94.6%
CRC012	89.3%	2.0%	91.3%
CRC020	90.2%	1.9%	92.1%
CRC021	68.0%	17.6%	85.6%
CRC026	90.1%	0.6%	90.7%
CRC027	92.1%	1.3%	93.4%
CRC034	74.3%	17.8%	92.1%
CRC035	86.4%	6.7%	93.1%
CRC036	91.7%	0.3%	92.0%
CRC040	84.6%	9.2%	93.9%
CRC047	76.6%	0.5%	77.2%
CRC052	75.3%	2.1%	77.3%
CRC065	87.9%	2.0%	89.9%
CRC098	79.8%	3.2%	83.0%
CRC099	78.7%	4.4%	83.0%
CRC101	79.2%	4.0%	83.3%
CRC102	89.8%	4.4%	94.2%
CRC106	68.8%	7.8%	76.6%
CRC108	89.2%	5.0%	94.2%
CRC114	86.4%	7.5%	93.9%
CRC125	89.1%	4.9%	94.1%
CRC138	89.9%	3.7%	93.5%

The remaining tasks for the proposal are:

Task 6: Development of the *k*-TSP classifier from mRNA-Seq (Months 18-24, Dr. Tan).

a. To develop predictive biomarkers from RNA-Seq, we will use the *k*-TSP algorithm in this proposal. Once we obtain the *P* x *N* matrix from the BiNGS! analysis, we will convert this into a rank-based matrix by ranking the expression of genes within a sample and perform standard normalization (where mean = 0 and standard deviation = 1). This relative rank-based matrix will be used as the training set for identifying predictive biomarkers.

b. We will use the S and R cell lines as previously defined as the training set to train the predictive classifier for an agent. Gene pairs with high scores are viewed as most informative for classification. Using an internal leave-one-out cross-validation, the final *k*-TSP classifier utilizes the *k disjoint* pairs of genes, which achieve the *k* best scores from the training set. In this study, the maximum number of pairs (kmax) is fixed at 10 to maintain feasibility for testing on clinical samples.

c. For human tumor explants, the *k*-TSP gene classifier will be performed on paraffin tissue blocks as previously described.

Task 7: Development of an integrated classifier (Months 18-24, Drs. Eckhardt and Tan). Since many studies have shown that ensemble approaches often outperform individual classifiers, integration of the *k*-TSP gene classifier with other molecular biomarkers such as gene sequencing and FISH data will be performed.

a. Gene mutation sequencing: for both CRC cell lines and human tumor explants, DNA will be isolated using the Qiagen DNA extraction kit (Qiagen, Valencia, CA). KRAS mutations will be analyzed by one of two methods. The human CRC explants will be assessed (in our CLIA-certified UCCC Pathology Core) using the DxS Scorpion method (DxS, Manchester, UK) according to the manufactures instructions. To avoid false-positive results due to background amplification, the assay will only be considered valid if the control Cp value is $\leq$ 35 cycles. Mutations will be scored positive when the DCp is less than the statistically set 5% confidence-value threshold. The CRC cell lines have been analyzed for KRAS mutations with a high resolution melting temperature method using custom primers and the Roche LC480 real time PCR machine (Mannheim, Germany). The additional CRC cell lines will also be assessed using this method. Other relevant gene mutations will be detected using previously published methods and primers.

b. Fluorescence in situ Hybridization (FISH): Dual-color FISH assays will be performed on the prepared slides of the CRC cell lines using 120 ng of Spectrum Red-labeled target probe (tailored for the agent) (UCCC Cytogenetics Lab) and 0.3 ml of a Spectrum Green-labeled centromeric probe (Abbott Molecular, Abbott Park, IL) per 113 mm^2 hybridization area according to previously published procedures. Analysis will be performed on an epifluorescence microscope using single interference filter sets for green (FITC), red (Texas Red), and blue (DAPI) as well as dual (red/green) and triple (blue, red, green) band

pass filters. Approximately 20 metaphase spreads and 100 interphase nuclei will be analyzed in each cell line, and ploidy assessed along with identification of the chromosomes harboring homologous sequences to the target/centrosome probe set. To determine occurrence of genomic imbalances, target copy number per cell will be compared to the expected by the ploidy of the cell line (e.g., 2 copies in diploid lines, 3 copies in triploid lines). For documentation, images are captured using a CCD camera and merged using dedicated software (CytoVision, AI, San Jose, CA).

c. The final prediction of this integrated classifier will be implemented as majority voting or weighted voting systems, depending on the training and validation data during the biomarker development step. Such a classifier can thus integrate both unbiased and biased biomarker discovery.

Task 8: Prioritization of agents to progress to Specific Aim 2 (Months 18-24, Drs. Eckhardt and Tan).

a. A classifier will be considered adequate to progress to Specific Aim 2 if it exhibits 90% accuracy against the validation set of cell lines (independent from the training set).

Aim 2. To validate the preclinical efficacy of these classifiers against 20 independent patient-derived CRC explant models.

Task 1: Prediction of the human CRC explants (Months 24-36, Drs. Eckhardt and Tan) The baseline human CRC explant will be assessed using RT-PCR for the *k*-TSP, FISH, and gene sequencing (see above for gene sequencing and FISH).

a. The *k*-TSP gene classifier will be performed on paraffin tissue blocks. The slides will be deparaffinized and then expression levels of genes in the TSP classifier will be assessed using RT-PCR. Total RNA will be isolated from cells using the RNeasy FFPE kit (Qiagen, Valencia, CA), cDNA synthesized from one microgram of total RNA using the Taqman reverse transcription kit (Applied Biosystems, Foster City, CA), and expression levels detected from 100 ng of cDNA using Power SYBR Green detection chemistry (Applied Biosystems, Foster City CA).

b. The prediction of the integrated classifier on the CRC explants will be implemented as majority voting or weighted voting systems, depending on the training and validation data during the biomarker development step.

Task 2: The human CRC explants will be treated with the agent and assessed for response (Months 24-36, Dr. Eckhardt).

a. See **Task 2a**. Obtaining tissue from CRC patients at the time of removal of a primary tumor or metastectomy is conducted under Colorado Multi-Institutional Review Board (COMIRB) approved protocols.

b. The relative tumor growth index (TGI) will be calculated by taking the relative tumor growth of treated mice divided by the relative tumor growth of control mice since the initiation of therapy (T/C) as described previously. Cases with a TGI of $\leq$50% will be considered sensitive; a TGI of >50% is considered resistant.

c. A classifier will be considered adequate to progress to clinical testing if it is 80% accurate against the 20 human tumor explants.

Final Data Analysis and Report Submission to the CDMRP: Months 35-36, Drs. Eckhardt and Tan.

Key Research Accomplishments:

- Completed in vitro screening on a large panel of CRC cell lines to determine the activity of six novel anti-cancer agents
- Completed baseline gene expression profiling of CRC cell lines and patient-derived tumor explants by high-throughput RNA-sequencing approach
- Analyzed the RNA-seq data with bioinformatics pipeline

Reportable outcomes: Based on the RNAseq data generated from this research, we have aligned our RNAseq data against the Cancer Genome Atlas (TCGA) colorectal cancer data. We have submitted an abstract on this topic that has been accepted for presentation at the 24th EORTC-NCI-AACR Symposium on Molecular Targets and Cancer Therapeutics, Dublin, Ireland (November 6-9, 2012).

Abstracts:

1. Tan AC, Britt BW, Astling DP, Leong S, Lieu C, Tentler JJ, Pitts TM, Arcaroli JJ, Messersmith WA, Eckhardt SG. (2012). Validation of Preclinical Colorectal Cancer Models Against TCGA Data for Pathway Analysis and Predictive Biomarker Discovery. To be Presented in the EORTC-NCI-AACR Symposium on Molecular Targets and Cancer Therapeutics, Dublin, Ireland.

Conclusions: We have completed **Task 1** within year 1, and continue to complete **Tasks 2** and **3** in the next 6 months. We have obtained high quality RNAseq data for colorectal cancer cell lines and tumor explants. The mappability of these RNAseq data against human reference genome >90%. We have completed **Tasks 4 and 5** within year 1. Our research plans for the next six months are to identify the preclinical models that are deemed extremely sensitive or resistant to the 6 anti-cancer agents *in vitro* and *in vivo*. These models will be used to train the predictive algorithm (**Task 6**). We will initiate the research efforts to identify potential mutations that correlate with sensitive to anti-cancer agents sensitivity, which we can incorporate to the development of the predictive classifiers (**Task 7**). We aim to identify the most promising anti-cancer agents by the end of Year 2 to move into **Aim 2** of this project.

REFERENCES:

Pitts TM, Tan AC, Kulikowski GN, Tentler JJ, Brown AM, Flanigan SA, Leong S, Coldren CD, Hirsch FR, Varella-Garcia M, Korch C, Eckhardt SG. (2010). Development of an integrated genomic classifier for a novel agent in colorectal cancer: approach to individualized therapy in early development. Clin Cancer Res. 16(12):3193-3204.

Skehan P, Storeng R, Scudiero D, Monks A, McMahon J, Vistica D, Warren JT, Bokesch H, Kenney S, Boyd MR. (1990). New colorimetric cytotoxicity assay for anticancer-drug screening. J Natl Cancer Inst. 82(13):1107-1112.

Trapnell, C., Pachter, L., and Salzberg, S.L. 2009. TopHat: discovering splice junctions with RNAseq. Bioinformatics 25:1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. 2010. Transcript assembly and quantification by RNAseq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511-515.

Appendix:
Abstract To be Presented in the EORTC-NCI-AACR Symposium on Molecular Targets and Cancer Therapeutics, Dublin, Ireland.

Validation of Preclinical Colorectal Cancer Models Against TCGA Data for Pathway Analysis and Predictive Biomarker Discovery

Aik Choon Tan[1], Byron W. Britt[1], David P. Astling[1], Stephen Leong[1], Christopher Lieu[1], John J. Tentler[1], Todd M. Pitts[1], John J. Arcaroli[1], Wells A. Messersmith[1], S. Gail Eckhardt[1]
[1]Division of Medical Oncology, Department of Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Background: Preclinical models such as cancer cell lines and patient-derived tumor xenografts (PDTX) have been widely used in predictive biomarker development and pathway modeling in cancer research. However, it has not been clear to what extent these preclinical models reflect the molecular heterogeneity observed in clinical samples, while initiatives such as the TCGA provide an opportunity for comparison and validation.

Methods: We performed massively parallel mRNA sequencing (RNA-seq) on 25 PDTX and 60 CRC cell lines using the Illumina HiSeq2000 platform to characterize the transcriptome of these preclinical models. On average, 40 million single-end 100bp sequencing reads per sample were obtained. The RNA-seq reads were mapped against the human genome using Tophat (version 1.3.2). On average, 80% of the reads aligned to the human genome. Cufflinks (version 1.3.0) was used to assemble the transcripts using the RefSeq annotation as the guide. Gene-level expression was estimated by FPKM (fragments per kilobase of exon per million fragments mapped). We performed pathway analysis using PARADIGM. RNA-seq of 244 CRC patient tumors were downloaded from the TCGA website. Following rank-normalized, mean centered data normalization, hierarchical clustering was performed on the samples using gene-centric and pathway-centric approaches.

Results: To determine whether the preclinical models were representative of the variability observed in expression profiles from clinical samples, we compared RNA-seq gene expression data of the 25 PDTX and 60 CRC cell lines with 244 TCGA CRC patient tumors. From the unsupervised hierarchical clustering approach, CRC cell lines and PDTX clustered together with TCGA patient tumors. We also performed unsupervised hierarchical clustering based on PARADIGM inferred gene sets. In the pathway clustering analysis, the preclinical CRC models also clustered together with TCGA patient samples. Within each cluster, CRC preclinical models do response to particular class of targeted therapy, suggesting potential treatment strategies for the diverse CRC patient samples.

Conclusions: In this study, we performed a systematic comparison of our CRC preclinical models and TCGA patient samples using next-generation sequencing data. Clustering analysis indicates that our preclinical models are representative of all CRC patient clusters identified in TCGA database. These results indicate that these CRC preclinical models are representative of actual patient samples and may be useful in early drug development and predictive biomarker discovery.